

Extending data curation to interdisciplinary and highly collaborative research

The School of Informatics, Computing, and Engineering at Indiana University (IU) in collaboration with IU Libraries, University of Colorado (CU) Boulder Libraries and University Libraries at Virginia Tech (VT) requests \$382,128 from IMLS with a \$9,654 cost share to fund research that extends data curation to interdisciplinary and highly collaborative research (IHCR). The proposed project will examine several use cases, develop a model of support that includes workflows and professional assistance, and evaluate this model through implementations in several libraries. It will generate deep knowledge of data practices in interdisciplinary research, engage the LIS community in collaborative development and evaluation, and enhance the long-term sustainability of data and its infrastructures.

STATEMENT OF NATIONAL NEED

IHCR¹ is a recognized priority in addressing complex research problems and providing solutions to societal challenges. Such research, however, has higher demands on communication, learning, and trust, all of which can have implications for how data is collected, managed, and preserved². Academic libraries began to develop some services in support of such research³, but the increasing number of U.S. interdisciplinary collaborations and the unique challenges of working with IHCR data highlight a national imperative to develop support models aligned with researchers' needs that fully leverage both libraries' expertise and library-research partnerships to improve IHCR data curation, sharing, and preservation. **This project addresses this national need by generating deep knowledge of IHCR data practices and developing and testing curation models that can be widely adopted by academic libraries to create successful interdisciplinary curation services.**

By developing a theoretical understanding of IHCR data and practical models of embedded curation built on it, the project contributes to the *Research in Service to Practice* category. It covers the piloting and scaling phases of maturity, as we will test our approaches with several use cases, expand them to testing and peer review within the broader community, and identify ways to incorporate adaptations and feedback.

PROJECT DESIGN

The proposed project addresses the following research questions: RQ1) What are the current data curation practices in IHCR? RQ2) How do IHCR teams negotiate differences related to data curation? RQ3) What is the role of libraries and librarians in supporting cultures of interdisciplinarity? Project goals include creating a curation service model generalizable across libraries, empowering librarians to become partners in IHCR, increasing library capacity to support heterogeneous data, and improving inter-institutional communication and collaboration.

We will rely on theories of action research and frameworks of purposeful curation and human practices design⁴ and carry out the work in four stages:

1. **Assessment of current practices** in 12 use cases with which we have relationships. These cases offer several integration challenges, including disciplinary diversity, data and instrumentation diversity, and organizational diversity (e.g., team members include members of the public or the government). Assessment data will be collected through participant observation and the Delphi methods. We will also convene a workshop that will bring together interdisciplinary scholars and librarians and collect more assessment data.

¹ We define IHCR as research that integrates resources and expertise across disciplines and institutional settings.

² Parsons, M. A., Godøy, Ø., LeDrew, E. et al. (2011). A conceptual framework for managing very diverse data for complex, interdisciplinary science. *Journal of Information Science*, 37(6), 555–569.

³ Witt, M. (2009). Institutional repositories and research data curation in a distributed environment. *Lib Trends*, 57(2).

⁴ Palmer, C. L., Weber, N. M., Munoz, T., & Renear, A. H. (2013). Foundations of data curation: The pedagogy and practice of “purposeful work” with research data. *Archive Journal*, 3.

2. **Workflow development** for interdisciplinary data, from collection to storage, analysis, and publication. These workflows will cover automated and “human-in-the-loop” processes and will emphasize data documentation and flows between collaborators and repositories.

3. **Implementation study.** We will engage students from LIS schools to help the use case teams implement and test workflows. The information gathered will be used to create a generalizable data curation model that harnesses existing university services, infrastructure, and engagement mechanisms to support data curation.

4. **Collaborative evaluation.** The workflows and best practices will be disseminated in the Data Curation Network (DCN)⁵ for peer review and feedback. We will also collaborate with DCN and other networks, e.g., the Research Data Alliance, to promote and organize community engagement workshops that will provide training in interdisciplinary data curation and invite interested libraries to evaluate the model at their institutions. The use of these knowledge networks will improve the model, adapt it to multiple contexts, and strengthen the national impact of the project.

OUTCOMES AND NATIONAL IMPACT

As libraries address burgeoning data and the complexities associated with its curation, they need new and extended models to satisfy this growing need. The proposed work will provide a new theoretical framework as well as a series of practical steps (workflows) to understand and improve IHCR data curation and publication. Additionally, our research will offer a model for connecting cross-institutional suites of data services and infrastructures to interdisciplinary research processes. Through combined professional networks, we are well positioned to scale up this effort up and see through its adaptation, adoption, and expansion.

The project’s national impact will be demonstrated in: a) increased knowledge of IHCR data practices that will allow libraries to expand data services and promote best practices in data curation; b) practical guidance in developing services and connecting them to end user needs; and c) new collaborations between libraries, researchers, and data management professionals. The focus on workforce training and collaborative evaluation will result in greater impact than an individual library could achieve. Ultimately, the insights gained from this project will contribute to the maturing of data curation and sharing cultures and to the increased impact of IHCR.

KEY PERSONNEL

Our project team combines technical, social science, and professional expertise that is strongly positioned for success. PI Dr. Inna Kouper will lead the project and its assessment and implementation studies. Co-PIs Jamie Wittenberg (IU Libraries), Dr. Thea Lindquist and Andrew Johnson (CU Boulder Libraries) will lead curation and collaboration activities. Co-PI Esen Tuna (IU) will coordinate expertise of workflow analysts and developers across the institutions. Co-PI Andrea L. Ogier (VT Libraries) will oversee testing and student engagement activities.

BUDGET SUMMARY AND SCHEDULE OF COMPLETION

We estimate a project lasting 24 months (9/1/20 - 8/31/22). The first four will focus on data practices assessment. The workflow development and the implementation study will proceed for the next eight months in an iterative process. The second year will focus on collaborative evaluation and adoption.

The total budget request is \$382,128 [indirect costs of \$122,613 (58.5%) on direct costs of \$214,167] with \$9,654 cost share. Co-PI salaries and fringes are \$126,734. Hourly student salary is \$28,800. Travel for collaboration and dissemination is \$10,500. Participant support for workshops is \$49,920. Supplies and transcription are \$5,500. Partners (CU Boulder and VT) are subcontractors with total budget of \$72,000 to cover salary, fringe, travel and overhead.

⁵ A network of ten academic institutions that share diverse expertise in data curation and technologies.